EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Directorate F – Health and Food
Disease Prevention

ECICC - EUROPEAN COMMISSION INITIATIVE ON COLORECTAL CANCER

European guidelines on colorectal cancer prevention, screening and diagnosis:

Draft systematic review protocol on colorectal cancer screening tests

Purpose of the document: for stakeholders' consultation

Date: October 2024

European Commission, Via Enrico Fermi 2749, I-21027 Ispra (Varese) - Italy. Telephone: (39)0332-78-9111.
E-mail: JRC-CANCER-POLICY-SUPPORT@ec.europa.eu

General structure

Healthcare Question	**Which screening test(s) or their combinations should be used for early detection of colorectal cancer (CRC) in asymptomatic adults at average risk?**
Objective	To assess and compare the effect of CRC screening tests on prioritised people-important outcomes
Design	We will conduct a systematic review using methods developed by the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2022) and the Cochrane Rapid Reviews Methods Group (Garrity 2021). We will also adhere to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement for reporting the process (Page 2021).
Eligibility criteria	**Population:** Asymptomatic adults at average risk
	Intervention(s): CRC screening tests implemented with any of the following selected tests or their combinations: - Colonoscopy - Flexible sigmoidoscopy - Faecal immunochemical test (FIT) - Guaiac Faecal Occult Blood Test (gFOBT) (as indirect evidence for FIT) - DNA stool-based test
	Comparator: CRC screening test and no screening
	Outcome(s): – CRC mortality – CRC incidence – Incidence of late-stage CRC (defined as stage III or IV or Duke's C

	or D) – Quality of life – related outcomes – Adverse event: Major Bleeding – Adverse event: Perforation
	Study design: We will primarily include randomized controlled trials (RCTs). If the evidence from RCTs is of moderate certainty or lower, due to imprecision or indirectness, we will consider inclusion of prospective observational studies, such as cohorts (Cuello-Garcia 2022). We will also consider the inclusion of modelling studies for simulation of the additional benefit, when the evidence for the critical outcomes from empirical studies (RCT or observational studies) is of moderate certainty or lower, due to imprecision or indirectness. We will exclude abstracts not published as full-text articles in peer-review journals. We will only include studies published in English.
	Context: Colorectal cancer screening
Search strategy	**Sources:** We will use electronic algorithms with a combination of controlled vocabulary and search terms in the following databases: i) MEDLINE; ii) EMBASE, and; iii) Web of Science Core. We will adapt the search algorithms to the requirements of each database, and we will use validated filters to retrieve appropriate designs, as needed. We will also review references of included studies, previous systematic reviews, and consult experts to capture additional studies potential missed from our original search. We will report in appendices, the complete search algorithms designed for each database, the hits retrieved, and the reasons for the exclusion of studies at the full text review stage.
	Data management We will use EndNote web software to create a database for the management of the search results and deduplication of citations

Study selection, evidence appraisal, and synthesis	**Study selection**: After a calibration procedure to standardise judgement criteria between reviewers, we will screen the retrieved citations to identify potential eligible studies, based on title and abstract. Subsequently, two reviewers will assess independently in order to confirm the eligibility of the citation pre-selected at title and abstract, after assessing the full text of each study. A third reviewer will confirm the eligibility, reviewing each assessment at this stage. In case of doubt at any step, another reviewer will be involved, and consensus will be reached by discussion. We will report this process in a PRISMA flowchart; we will record the decision and conduct the calibration process using the Rayyan platform.
	Data collection: One reviewer will extract the main characteristics of included studies in a tabulated format, including: author and publication year, study design, description of screening modality, study period, description of participants (number, screening adherence, age and sex distribution), adherence, length of follow-up and outcomes data (events and numbers of patients included for analyses in each group), and conflict of interest. A second reviewer will check the extracted data for accuracy, and any disagreement will be solved by consensus or involving a third author.
	Risk of bias (outcome level): We will evaluate the risk of bias for each outcome using tools specifically designed for each type of study design.
	Strategy for data synthesis: We will report the estimates stratified by type of study design for each outcome.
	Analysis of subgroups or subsets: We will report the outcomes' estimates according to the following predefined subgroups: - Periodicity of screening - Age of screening / age where the screening starts (before 50 and 50+) - Sex (female and male) - Cancer localization (proximal or distal)

	We will conduct sensitivity analyses restricted to studies presenting overall low risk of bias.
Summary of findings, and assessment of the certainty of evidence	We will rate the certainty of evidence across design of studies and for each outcome following the GRADE approach as high, moderate, low or very low, depending on several factors including risk of bias, imprecision, inconsistency, indirectness and publication bias (Schünemann 2013).
References	– Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook. – Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, Affengruber L, Stevens A. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol. 2021 Feb;130:13-22 – Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160. – Cuello-Garcia CA, Santesso N, Morgan RL, Verbeek J, Thayer K, Ansari MT, Meerpohl J, Schwingshackl L, Katikireddi SV, Brozek JL, Reeves B, Murad MH, Falavigna M, Mustafa R, Regidor DL, Alexander PE, Garner P, Akl EA, Guyatt G, Schünemann HJ. GRADE guidance 24 optimizing the integration of randomized and non-randomized studies of interventions in evidence syntheses and health guidelines. J Clin Epidemiol. 2022 Feb;142:200-208 – Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. Available from: www.guidelinedevelopment.org/handbook.

Annex 1. Methods specific for empirical evidence (Randomized controlled trials and observational studies)

Study selection, data collection, and risk of bias.	**Study selection**: The selection process is reported in the General structure section. For empiric evidence, we will be flexible in the age range of screening. We will include only the comparison of the following screening periodicities: 1) annual or biennial for FIT and gFOBT; 2) once in a lifetime, five, ten, and fifteen years for colonoscopy and sigmoidoscopy; and 3) annual, biennial, triennial, and five years for DNA stool-based tests.
	Data collection: The data collection is reported in the General structure section.
	Risk of bias (outcome level): One reviewer will assess the risk of bias of included studies, using Version 2 of the Cochrane risk-of-bias tool for randomised trials (RoB 2). The tool is structured in a set of bias domains, focussing on different aspects of trial design, conduct, and reporting. The five domains for individually randomised trials are: 1. Bias arising from the randomization process. 2. Bias due to deviations from intended interventions. 3. Bias due to missing outcome data 4. Bias in measurement of the outcome 5. Bias in selection of the reported result. For observational studies, we will use the Risk of Bias in Non-Randomized Studies of Interventions (ROBINS-I) tool, which consist on seven domains: 1. confounding 2. selection of participants into the study, address issues before the start of the interventions that are to be compared ("baseline"). 3. classification of the interventions 4. biases due to deviations from intended interventions, 5. missing data, 6. measurement of outcomes, and selection of the reported result. Both tools include algorithms that map responses to signalling questions onto a proposed risk-of-bias judgement for each domain (low, some concerns, or high). The response options for overall risk-of-bias judgement are the same as for individual domains. A second reviewer will confirm the assessment. In case of doubt, another reviewer will be involved and consensus will be reached by discussion.

Data synthesis	**Strategy for data synthesis:** We will report the estimates stratified by the type of study design for each outcome for which we decided to include other types of studies, apart from RCTs. The evidence from RCTs will be focused on efficacy outcomes, while harms will be evaluated using observational studies. For the comparison between tests, we will consider synthesizing the evidence for a base case scenario. This scenario is defined as the most common strategies and periodicities used in the European Union (e.g. biennial periodicity for stool-based tests and a ten-year periodicity for endoscopy tests). Randomized controlled trials We will synthesize the evidence from RCTs through a network meta-analysis. We will assess the transitivity assumption of NMA comparing the distribution of main prognostic factors (sex, age) across RCTs. We will graphically represent the network of comparisons for efficacy outcomes, through a network map including all screening studies included, regardless of comparison (Chaimani 2013). We will estimate the relative effects of the screening tests, pooling effect sizes (i.e. relative risks) for each efficacy (i.e. mortality) outcome with multivariate random-effects meta-analysis under the consistency contrast-based model and a frequentist framework. We will assume that the between-study heterogeneity variance is common to all comparisons or treatment contrasts (Higgins 2012). We will assess the consistency of the network. If there is at least one closed loop in the network, we will assess the consistency (statistical manifestation of transitivity) of the network through both local and global approaches. Global inconsistency will be assessed with the design-by-treatment interaction inconsistency model of meta-analysis (Higgins 2012). If inconsistency is detected, we will employ the loop-specific and the node-splitting methods to identify the source of the inconsistency (Higgins 2012). Identification of local inconsistency will also be conducted by assessing the differences in point estimates and the overlap in the confidence intervals from direct and indirect estimates. We will perform all analyses in Stata, using the network suite of commands (White 2015). Observational studies We will synthesize the evidence from observational studies through a pairwise meta-analysis. We will estimate the overall effectiveness of screening prioritizing the adjusted estimations. We will pool effect sizes (i.e. relative risks, hazard ratios) for efficacy (i.e. mortality) using a random effects model with the

	Mantel-Haenzel or inverse variance method. To estimate between-study variance and confidence intervals, we used the Paule-Mendel and Q-profile methods. For harms, we will pool proportions (events over exposed subjects to intervention) implementing a generalised linear mixed random model with a logit transformation, and the Clopper-Pearson method to estimate the confidence interval for individual study results. We will assess the presence of heterogeneity between studies by visual inspection of forest plots for all outcomes and complemented by the assessment of the *Q* statistic and I^2 parameters for relative effects (Rücker 2008). Also for relative effects heterogeneity of 0% to 40% will be considered as "might not be important," 30% to 60% as "moderate heterogeneity," 50% to 90% as "substantial heterogeneity," and 75% to 100% as "considerable heterogeneity." Noteworthy, overlapping categories convey that there are no strict cut-offs for interpreting heterogeneity, and categorization depends on the magnitude and direction of effects We will perform all analysis in RStudio.
	Analysis of subgroups or subsets: The subgroup analysis is reported in the General structure section.
Summary of findings, and assessment of the certainty of evidence	We will rate the certainty of evidence across studies and for each outcome following the GRADE approach as high, moderate, low or very low, depending on several factors including risk of bias, imprecision, inconsistency, indirectness and publication bias (Schünemann 2013). For RCT evidence, appropriate methods for assessing certainty of evidence from a network metanalysis will be followed (Izcovich 2023). To assess the domain of imprecision, we will set thresholds for rating down our certainty, based on the minimal important difference for each outcome (Schünemann 2022). The thresholds are determined based on the median responses obtained from a survey conducted by the working group (See JRC technical report on CRC screening). The thresholds will be categorised into two levels for each outcome: 1) Colorectal Cancer mortality: small/moderate= 95 per 100,000; and moderate/large=175 per 100,000; 2) Diagnosis of Colorectal Cancer: small/moderate= 200 per 100,000; and moderate/large=375 per 100,000; 3) Major bleeding: small/moderate= 550 per 100,000; and moderate/large=950 per 100,000; 4) Colonic perforation: small/moderate= 450 per 100,000; and moderate/large=775 per 100,000. We will develop GRADE evidence profiles and summary of findings (SoF) tables, summarising the evidence for a list of selected outcomes, the

	direct, indirect and network summary relative effects (only for RCT evidence), absolute effects of the intervention and the certainty of evidence (Yepes-Nuñez 2019).
References	• Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PLoS One. 2013 Oct 3;8(10):e76654. • Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods. 2012 Jun;3(2):98-110. • Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook. • Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. BMJ. 2023 Jun 27;381:e074495. doi: 10.1136/bmj-2022-074495. PMID: 37369385. • White IR. Network meta-analysis. Stata Journal 2015; 15: 951–985. • Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC medical research methodology. 2008 Dec;8(1):79 • Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. Available from: www.guidelinedevelopment.org/handbook. • Schünemann HJ, Neumann I, Hultcrantz M, Brignardello-Petersen R, Zeng L, Murad MH, Izcovich A, Morgano GP, Baldeh T, Santesso N, Cuello CG, Mbuagbaw L, Guyatt G, Wiercioch W, Piggott T, De Beer H, Vinceti M, Mathioudakis AG, Mayer MG, Mustafa R, Filippini T, Iorio A, Nieuwlaat R, Marcucci M, Coello PA, Bonovas S, Piovani D, Tomlinson G, Akl EA; GRADE Working Group. GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. J Clin Epidemiol. 2022 Oct;150:225-242 • Yepes-Nuñez JJ, Li SA, Guyatt G, Jack SM, Brozek JL, Beyene J, Murad MH, Rochwerg B, Mbuagbaw L, Zhang Y, Flórez ID, Siemieniuk RA, Sadeghirad B, Mustafa R, Santesso N, Schünemann HJ. Development of the summary of findings table for network meta-analysis. J Clin Epidemiol. 2019 Nov;115:1-13. doi: 10.1016/j.jclinepi.2019.04.018. Epub 2019 May 2. PMID: 31055177.

Annex 2. Methods specific for modelling evidence

Study selection, data collection, and risk of bias.	**Study selection**: The selection process is reported in the General structure section. We will only include microsimulation modelling studies with a lifetime horizon, that evaluate the comparison of the following screening periodicities: 1) annual or biennial for FIT and gFOBT; 2) once in a lifetime, five, ten, and fifteen years for colonoscopy and sigmoidoscopy; and 3) annual, biennial, triennial, and five years for DNA stool-based tests. Since a modelling study could simulate every screening scenario, we will restrict only the modelling studies that evaluated any of the prioritized screening strategies, starting at 45 or 50 and end at 69, 74, or 79 years old, or their equivalent in the number of rounds of each screening strategy and periodicity, with an exception of once-in-a-lifetime endoscopy test. For once-in-a-lifetime strategy, we will not restrict the studies considering the age of start.
	Data collection: The data collection is reported in the General structure section.
	Risk of bias (outcome level): Two reviewers will assess independently the credibility (quality) of the model using the ISPOR/ISMD checklist, which includes signalling question for separate domains assessing relevance (directness of the model) and credibility (Jaime Caro 2014). In case of doubt another reviewer will be involved and consensus will be reached by discussion.
Data synthesis	**Strategy for data synthesis:** For the comparison between tests, we will consider synthesizing the evidence for a base case scenario. This scenario is defined as the most common strategies and periodicities used in the European Union (e.g. biennial periodicity for stool-based tests and a ten-year periodicity for endoscopy tests). We will estimate the absolute number of events per outcome, by subtracting the effect estimated in the interventions from the comparator for each study. If more than one study provides estimations of the outcome, we will describe the median of the absolute effect and the range of effect per 1000 persons.

	Analysis of subgroups or subsets: The subgroup analysis is reported in the "general structure".
Summary of findings, and assessment of the certainty of evidence	We will rate the certainty of evidence across studies and for each outcome following the GRADE approach as high, moderate, low or very low, depending on several factors. For evidence from modelling studies we will apply the GRADE approach guidance to assess the evidence for this type of studies which considers both the certainty of inputs parameter and credibility of the model along the other domains (Brozek 2021). We will develop GRADE evidence profiles and summary of findings (SoF) tables, summarising the evidence for a list of selected outcomes, the direct, indirect and network summary relative effects, absolute effects of the intervention and the certainty of evidence.
References	• Jaime Caro J, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, Briggs AH; ISPOR-AMCP-NPC Modeling CER Task Forces. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014 Mar;17(2):174-82 • Brozek JL, Canelo-Aybar C, Akl EA, Bowen JM, Bucher J, Chiu WA, Cronin M, Djulbegovic B, Falavigna M, Guyatt GH, Gordon AA, Hilton Boon M, Hutubessy RCW, Joore MA, Katikireddi V, LaKind J, Langendam M, Manja V, Magnuson K, Mathioudakis AG, Meerpohl J, Mertz D, Mezencev R, Morgan R, Morgano GP, Mustafa R, O'Flaherty M, Patlewicz G, Riva JJ, Posso M, Rooney A, Schlosser PM, Schwartz L, Shemilt I, Tarride JE, Thayer KA, Tsaioun K, Vale L, Wambaugh J, Wignall J, Williams A, Xie F, Zhang Y, Schünemann HJ; GRADE Working Group. GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence-An overview in the context of health decision-making. J Clin Epidemiol. 2021 Jan;129:138-150. doi: 10.1016/j.jclinepi.2020.09.018.